

# EUCLIDEAN DISTANCE BETWEEN HAAR ORTHOGONAL AND GAUSSIAN MATRICES

C. E. GONZÁLEZ-GUILLÉN, C. PALAZUELOS, AND I. VILLANUEVA

**ABSTRACT.** In this work we study a version of the general question of how well a Haar distributed orthogonal matrix can be approximated by a random Gaussian matrix. Here, we consider a Gaussian random matrix  $Y_n$  of order  $n$  and apply to it the Gram-Schmidt orthonormalization procedure by columns to obtain a Haar distributed orthogonal matrix  $U_n$ . If  $F_i^m$  denotes the vector formed by the first  $m$ -coordinates of the  $i$ th row of  $Y_n - \sqrt{n}U_n$  and  $\alpha = \frac{m}{n}$ , our main result shows that the Euclidean norm of  $F_i^m$  converges exponentially fast to  $\sqrt{\left(2 - \frac{4}{3} \frac{(1-(1-\alpha)^{3/2})}{\alpha}\right)} m$ , up to negligible terms.

To show the extent of this result, we use it to study the convergence of the supremum norm  $\epsilon_n(m) = \sup_{1 \leq i \leq n, 1 \leq j \leq m} |y_{i,j} - \sqrt{n}u_{i,j}|$  and we find a coupling that improves by a factor  $\sqrt{2}$  the recently proved best known upper bound of  $\epsilon_n(m)$ . Our main result has also applications in Quantum Information Theory.

## 1. INTRODUCTION

One of the classical problems in random matrix theory is to compare a random Gaussian matrix  $Y_n = (y_{i,j})_{i,j=1}^n$  with a Haar distributed random matrix  $U_n = (u_{i,j})_{i,j=1}^n$  in the orthogonal group  $\mathcal{O}(n)$ . This problem dates back to the famous work [2], where Borel showed that the distribution of one single coordinate of  $U_n$  converges to the distribution of one single coordinate of  $Y_n$ , when properly normalized. That is, for a fixed pair  $(i, j)$  we have that  $\sqrt{n}u_{i,j}$  converges in distribution to a standard normal. Since then, many authors have studied the problem of how many entries of  $\sqrt{n}U_n$  can be simultaneously well approximated by the corresponding entries of  $Y_n$ ; that is, by independent standard normal distributions. Indeed, a number of papers in the 1980's made further progress in this direction (see for instance [10], [11], [18], [20]). The reader can find a more complete introduction to this problem in [8, Section 6.3] and the references therein.

Some of the previous works focused on the variation distance. More precisely, in [18] the authors proved that the joint distribution of the first  $l_n$  coordinates of the first column of  $Y_n - \sqrt{n}U_n$  converges to 0 in variation distance as  $n$  grows to

---

The final publication is available at Springer via <http://dx.doi.org/10.1007/s10959-016-0712-6>

infinity, provided that  $l_n = o(\sqrt{n})$ . Later, the order was improved to  $l_n = o(n)$  ([10]). More recently, in [9], it was proven that the joint distribution of the upper left  $l_n \times m_n$  block of  $Y_n - \sqrt{n}U_n$  converges to 0 in variation distance provided that  $l_n, m_n$  are both  $o(n^{\frac{1}{3}})$ . Later, in [3], this order was improved to  $O(n^{\frac{1}{3}})$ .

The latest major achievement in this direction came from [14, 15]. In those papers the author showed that the joint distribution of the upper left  $l_n \times m_n$  block of  $Y_n - \sqrt{n}U_n$  converges to 0 in variation distance if and only if  $l_n, m_n$  are both  $o(n^{\frac{1}{2}})$ . This settled the long standing open problem of finding the best ratio in the variation distance case (see [8, Section 6.3]). In the same paper Jiang also showed [14, Theorem 3] the existence of a coupling between  $Y_n$  and  $U_n$  such that

$$\epsilon_n(m) = \sup_{1 \leq i \leq n, 1 \leq j \leq m} |y_{i,j} - \sqrt{n}u_{i,j}|$$

converges to 0 in probability if and only if  $m = o(\frac{n}{\log n})$ . Moreover, if  $m = \frac{\beta n}{\log n}$  then the previous supremum converges in probability to  $2\sqrt{\beta}$ . These results have been applied in [19] to study the eigenvector distribution of a wide class of Wigner ensembles. For further history and applications of these results, see [14, 15, 19]. Recent different generalizations of Borel's theorem can be seen in [4, 5].

Given the relevance of the Euclidean norm in many contexts, and motivated by these previous works, in this paper we study the behavior of the Euclidean norm of blocks of  $Y_n - \sqrt{n}U_n$ . We are interested not only in the order needed for convergence to 0, but in the general value of the norm. To show the extent of our main result, we show later how to recover from it one of the main results of [14]. We have also applied it to solve a question in Quantum Information Theory [12].

Let us define the notation needed to state our main result. Our probability spaces will be  $\mathbb{R}^{n^2}$  with the Gaussian measure. For every  $n \in \mathbb{N}$ ,  $Y_n = (y_{i,j})_{i,j=1}^n$  will be a Gaussian random matrix; that is, the variables  $(y_{i,j})_{i,j=1}^n$  are independent standard normal variables. For every  $1 \leq j \leq n$  we consider the column vector  $\mathbf{y}_j = (y_{i,j})_{i=1}^n$ . With probability 1, they form a basis of  $\mathbb{R}^n$ . Following [14], we apply the Gram-Schmidt orthonormalization procedure to  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$  to obtain an orthonormal basis  $(\boldsymbol{\nu}_j)_{j=1}^n$ . We call  $U_n$  the matrix  $(\nu_{i,j})_{i,j=1}^n$ . We recall that  $U_n$  is Haar distributed.

For every  $1 \leq m \leq n$  and for every  $1 \leq i \leq n$ , let  $F_i^m$  be the vector formed by the first  $m$ -coordinates of the  $i$ th row of  $Y_n - \sqrt{n}U_n$ . We describe the asymptotic generic behavior of  $\|F_i^m\|$ , where  $\|\cdot\|$  is the Euclidean norm. Let  $[x]$  denote the integer part of  $x$ .

**Theorem 1.1.** *Let  $n \in \mathbb{N}$ , let  $0 < \alpha \leq 1$  be fixed and let  $m = [\alpha n]$ . Let  $Y_n, U_n, F_i^m$  be as above. Then, there exists  $0 < \delta < \frac{1}{2}$  such that,*

$$\sup_i \|F_i^m\| \leq \sqrt{\left(2 - \frac{4}{3} \frac{(1 - (1 - \alpha)^{3/2})}{\alpha}\right) m} + O(m^\delta)$$

and

$$\inf_i \|F_i^m\| \geq \sqrt{\left(2 - \frac{4}{3} \frac{(1 - (1 - \alpha)^{3/2})}{\alpha}\right) m} - O(m^\delta),$$

both with probability exponentially close to 1 as  $n$  grows to infinity.

We have chosen this presentation of the main theorem for the sake of clarity. The actual proof shows further insight into the result. Specifically, we want to mention that there is a trade off between the rate of the concentration and the order  $\delta$  appearing in Theorem 1.1. In our proof we show how to make  $\delta = \frac{2}{5}$  keeping a very fast concentration rate. Nevertheless, the parameters can be changed easily to obtain a different value for  $\delta$ , at the cost of modifying the rate of the exponential convergence of the probability.

We clarify next some aspects of our result. The coupling is given by the Gram-Schmidt procedure performed columnwise. It is simple to see that the expectation of the norms of the columns of  $Y_n - \sqrt{n}U_n$  increases with the index of the column. Our main result says that the norms of the truncated rows will all concentrate exponentially around the same value. This “flatness” phenomenon is very relevant in our applications.

Therefore, our main contribution can be seen as a “block-delocalization” result. We show that the Euclidean norm of the whole block

$$\left( \sum_{1 \leq i \leq n, 1 \leq j \leq m} |y_{i,j} - \sqrt{n}u_{i,j}|^2 \right)^{\frac{1}{2}}$$

is well delocalized among the Euclidean norm of the rows. The lack of independence is the main difficulty in this case and we need to deal with different technicalities to overcome this and prove our result. Our main tools are standard versions of the concentration of measure phenomenon and the Gram-Schmidt procedure.

Theorem 1.1 considers the case of constant  $\alpha = \frac{m}{n}$ . We describe in the following corollary the behaviour of the euclidean norm of the rows in the case  $\alpha \rightarrow 0$ .

**Corollary 1.2.** *For every  $n \in \mathbb{N}$ , choose  $m \in \mathbb{N}$  with  $0 \leq m \leq n$  so that the sequence  $\alpha_n = \frac{m}{n}$  verifies  $\lim_{n \rightarrow \infty} \alpha_n = 0$ . Let  $Y_n, U_n, F_i^m$  be as above. Then,*

$$\sup_i \|F_i^m\| \leq \frac{m}{\sqrt{2n}} + o\left(\frac{m}{\sqrt{n}}\right),$$

and, similarly,

$$\inf_i \|F_i^m\| \geq \frac{m}{\sqrt{2n}} - o\left(\frac{m}{\sqrt{n}}\right),$$

both with probability greater than  $1 - Cne^{-m^k}$  for certain positive constants  $k$  and  $C$ .

In particular,  $\lim_{n \rightarrow \infty} \sup_i \|F_i^m\| \rightarrow 0$  if and only if  $m = o(\sqrt{n})$ .

As a consequence of Theorem 1.1 we prove in Section 4 a result about the supremum norm  $\epsilon_n(m)$ .

**Theorem 1.3.** *For each  $n \geq 2$ , there exist matrices  $Y'_n = (y'_{ij})_{i,j=1}^n$  and  $U'_n = (u'_{i,j})_{i,j=1}^n$  whose  $2n^2$  entries are real random variables defined on the same probability space such that*

- (i) *the law of  $U'_n$  is the normalized Haar measure on the orthogonal group  $\mathcal{O}(n)$ ;*
- (ii)  *$\{y'_{i,j}; 1 \leq i, j \leq n\}$  are independent standard normals;*
- (iii) *set*

$$\epsilon_n(m) = \max_{1 \leq i \leq n, 1 \leq j \leq m} |\sqrt{n}u'_{i,j} - y'_{i,j}|$$

*for  $m = 1, 2, \dots, n$ . Then, there exists  $0 < \delta < \frac{1}{2}$  such that for any  $\varepsilon > 0$  we have*

$$\begin{aligned} \epsilon_n(m) &\geq (1 - \varepsilon)(\sqrt{\varphi(\alpha)} - O(m^{-\delta}))\sqrt{2 \log n} \quad \text{and} \\ \epsilon_n(m) &\leq (1 + \varepsilon)(\sqrt{\varphi(\alpha)} + O(m^{-\delta}))\sqrt{2 \log(nm)} \end{aligned}$$

*with probability  $1 - o(1)$ , where we consider  $0 < \alpha \leq 1$  fixed,  $m = \lfloor \alpha n \rfloor$  and  $\varphi(\alpha) = 2 - \frac{4}{3} \frac{(1-(1-\alpha)^{3/2})}{\alpha}$  is the function appearing in Theorem 1.1.*

If we let  $\alpha$  change with  $n$  in Theorem 1.3 so that  $m_n = o(\frac{n}{\log n})$  we recover the convergence to 0 already obtained in [14, Theorem 3] (see Corollary 4.2). Furthermore, if we pick  $m_n = \frac{\beta n}{\log n}$  we get that

$$\sqrt{\beta} \leq \epsilon_n(m) \leq \sqrt{2\beta}.$$

Note that in [14, Theorem 3] the author obtained for this case  $\epsilon_n(m) \rightarrow 2\sqrt{\beta}$ . Therefore, we improve the upper bound by a factor  $\sqrt{2}$ . The key point is that our Theorem 1.1 allows us to modify the coupling. The price we pay is that we do not obtain an explicit coupling, but a randomized one (with high probability). Also, we have not been able to show that this bound is tight. Details are given in Section 4.

The fact that Theorem 1.3 follows from Theorem 1.1 provides a better understanding of the order  $\frac{n}{\log n}$  needed for the convergence of the supremum norm of the

block. Roughly, each of the row vectors of the difference, when multiplied from the right by a random unitary, distributes uniformly on the unit sphere. Therefore, the distance between its supremum and Euclidean norms is of the order  $\log m$ .

One of our original motivations to study this problem was to solve a question in Quantum Information Theory. In that context, one is interested in the matrices  $\Gamma = G_1 G_2^T$ , where  $G_1$  and  $G_2$  are two independent  $n \times m$  real Gaussian matrices, whose rows are normalized in the Euclidean norm. This definition of  $\Gamma$  guarantees that it belongs to the set of the so called *quantum correlation matrices*. An interesting question in the context of Bell inequalities is to study whether these matrices belong to the set of *local correlation matrices*, which can be mathematically expressed by the fact that  $\Gamma$  belongs to the unit ball of  $\ell_\infty^n \otimes_\pi \ell_\infty^n$ , where  $\pi$  denotes the projective tensor norm. The main result in [12] states that for a certain constant  $\alpha_0 \in (0, 1)$  one has that  $\Gamma$  is nonlocal with probability tending to one as  $n$  tends to infinity, provided that  $\frac{m}{n} \leq \alpha_0$ . The key point in the proof is that a similar result can be obtained from the recent work [1] if one replaces the Gaussian matrices  $G_1$  and  $G_2$  above with two truncated matrices independently distributed in the orthogonal group of order  $n$ . Hence, Theorem 1.1 is the right “tool” to pass from orthogonal to Gaussian matrices, at the cost of a slight modification in the constant  $\alpha_0$ . We refer to [12] for details.

The rest of the paper is organized as follows. In Section 2 we define our notation and we recall several known facts about the Gaussian distribution that will be repeatedly used later on. Then, in Section 3 we prove our main result Theorem 1.1. At the end of that section we show how simple modifications of the same techniques prove Corollary 1.2. In Section 4 we apply Theorem 1.1 to the study of the supremum norm of the  $n \times m$  blocks of  $Y_n - \sqrt{n}U_n$  and we prove Theorem 1.3. Next, we obtain as a corollary a slight improvement of [14, Theorem 3].

## 2. PRELIMINARIES

In this section we define our notation and for the sake of completeness we recall several known facts about the Gaussian measure on  $\mathbb{R}^n$  that will be used several times in the rest of the paper. We say that a real function  $f(n)$  is  $O(g(n))$  if there exist constants  $C > 0$  and  $n_0$  such that for all  $n > n_0$  we have that  $|f(n)| \leq Cg(n)$ . We say that  $f(n)$  is  $o(g(n))$  if  $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$ . We will use  $O(g(x))$  and  $o(g(x))$  to denote functions on these sets. We will say that a sequence of events  $E_n$  holds with probability exponentially small (respectively exponentially close to 1) if there exists  $\alpha > 0$ , independent of  $n$  such that  $Pr(E_n) \leq O(e^{-n^\alpha})$  (respectively  $Pr(E_n) \geq 1 - O(e^{-n^\alpha})$ ).

We recall the following well known bounds of the tail of a normal random variable.

**Lemma 2.1.** *Let  $Z$  be a standard normal random variable. Then, for every  $t > 0$ ,*

$$\frac{t}{(1+t^2)\sqrt{2\pi}}e^{-\frac{t^2}{2}} \leq \Pr(Z > t) \leq \frac{1}{t\sqrt{2\pi}}e^{-\frac{t^2}{2}}.$$

Hence, for  $t \geq 1$ ,

$$\Pr(Z^2 > t^2) \leq e^{-\frac{t^2}{2}}.$$

We will later choose  $t = m^{\frac{\epsilon}{2}}$  to get

$$\Pr(Z^2 > m^\epsilon) \leq e^{-\frac{m^\epsilon}{2}}.$$

We will denote the standard Gaussian probability measure (Gaussian measure in short) in  $\mathbb{R}^n$  by  $\mathcal{G}_n$ . We will refer to a Gaussian vector (matrix) as a random vector whose coordinates are independent standard normal random variables in  $\mathbb{R}$ .

The following bound of the norm of a Gaussian vector is well known. It can be easily deduced, for instance, from [16, Lemma 1].

**Proposition 2.2.** *For every  $0 < \epsilon < 1$ ,*

$$\mathcal{G}_n\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \geq \frac{\sqrt{n}}{\sqrt{1-\epsilon}}\} \leq e^{-\frac{\epsilon^2 n}{4}}$$

and

$$\mathcal{G}_n\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq \sqrt{n}\sqrt{1-\epsilon}\} \leq e^{-\frac{\epsilon^2 n}{4}}.$$

We will use several times along the paper the well known fact that both the Gaussian measure  $\mathcal{G}_n$  in the space of vectors  $\mathbb{R}^n$  and the Gaussian measure  $\mathcal{G}_{n^2}$  in the space of square matrices of order  $n$  are biunitarily invariant under the action of the orthogonal group  $\mathcal{O}(n)$ . Using this, it is very easy to see that the projection  $P_L(\mathbf{x})$  of a random Gaussian vector  $\mathbf{x}$  onto a fixed subspace  $L$  of dimension  $k$  is a Gaussian vector of this subspace.

One can see the rotationally invariant (uniform) measure  $\mu_n$  in  $S^{n-1}$  as the pushforward measure of  $\mathcal{G}_n$  given by the map  $f(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ . That is, given a set  $A \subset S^{n-1}$  we have that  $\mu_n(A) = \mathcal{G}_n(f^{-1}(A))$ .

Similarly, one can consider the pushforward measure of  $\mathcal{G}_{n^2}$  induced by the map that takes the first  $k$   $n$ -dimensional vectors  $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$  to the  $\text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ , the linear subspace generated by them. This measure is the only one invariant under the action of  $\mathcal{O}(n)$  and therefore we call it the Haar measure in the Grassmannian of the  $k$ -dimensional subspaces of  $\mathbb{R}^n$ .

The following proposition follows immediately from the previous explanation.

**Proposition 2.3.** *Let  $L \subset \mathbb{R}^n$  be a Haar distributed  $k$ -dimensional subspace and let  $\mathbf{x} \in \mathbb{R}^n$  be a Gaussian vector independent from  $L$ . Then, for any  $0 < \epsilon < 1$ ,*

$$Pr \left( \|P_L(\mathbf{x})\|_2 \geq \frac{\sqrt{k}}{\sqrt{1-\epsilon}} \right) \leq e^{-\frac{\epsilon^2 k}{4}}$$

and

$$Pr \left( \|P_L(\mathbf{x})\|_2 \leq \sqrt{k}\sqrt{1-\epsilon} \right) \leq e^{-\frac{\epsilon^2 k}{4}}.$$

If we replace the Gaussian vector by a fixed unitary vector we obtain the following estimates.

**Proposition 2.4.** *Let  $L \subset \mathbb{R}^n$  be a Haar distributed  $k$ -dimensional subspace and let  $\mathbf{y} \in \mathbb{R}^n$  be a fixed unitary vector. Then, for any  $0 < \rho < 1$  we have*

$$Pr \left( \|P_L(\mathbf{y})\| \geq \frac{1}{1-\rho} \sqrt{\frac{k}{n}} \right) \leq e^{-\frac{\rho^2 k}{4}},$$

and

$$Pr \left( \|P_L(\mathbf{y})\| \leq (1-\rho) \sqrt{\frac{k}{n}} \right) \leq e^{-\frac{\rho^2 k}{4}}.$$

For  $t > 1$  we also have

$$Pr \left( \|P_L(\mathbf{y})\| \geq t \sqrt{\frac{k}{n}} \right) \leq e^{-\frac{k}{4}(t^2-2)}.$$

*Proof.* One can consider a Haar distributed  $k$ -dimensional subspace  $L$  as a Haar distributed orthogonal matrix  $U$  acting on a fixed  $k$ -dimensional subspace  $M$ . Hence,  $P_L(\mathbf{y}) = P_M(U\mathbf{y})$ . Now, the vector  $U\mathbf{y}$  is a random uniform vector on the unit sphere of  $\mathbb{R}^n$  and, according to our explanation above, it is  $\bar{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$  for a Gaussian vector  $\mathbf{x}$ . Thus,  $P_L(\mathbf{y})$  has the same distribution as  $P_M(\bar{\mathbf{x}})$ . Then, the result can be easily deduced from the known estimates on  $P_M(\bar{\mathbf{x}})$ , for example, from [6, Lemma 2.2]. Also, note that a version of this proposition with slightly worse constants of the first two bounds can be easily deduced from Proposition 2.2 and Proposition 2.3.  $\square$

### 3. PROOF OF THEOREM 1.1

We briefly recall our notation:  $Y_n = (y_{i,j})_{i,j=1}^n$  will be a normal Gaussian random matrix. We consider the column vectors  $\mathbf{y}_j = (y_{i,j})_{i=1}^n$ . With probability 1, they form a basis of  $\mathbb{R}^n$  and, in that case, we can apply the Gram-Schmidt orthonormalization procedure to  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$  and we obtain an orthonormal basis  $(\boldsymbol{\nu}_j)_{j=1}^n$ . We call  $U_n$  the matrix  $(\nu_{i,j})_{i,j=1}^n$ . For every  $1 \leq m \leq n$  and for every  $1 \leq i \leq n$ ,  $F_i^m$  is the vector formed by the the first  $m$ -coordinates of the  $i$ th row of  $Y_n - \sqrt{n}U_n$ .

We start the proof of Theorem 1.1 with some observations about the Gram-Schmidt orthonormalization process. Let us examine the situation in step  $j$ . The Gaussian vectors  $\mathbf{y}_1, \dots, \mathbf{y}_{j-1}$  have been chosen independently. Associated to them we have the orthonormal vectors  $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{j-1}$ . Both sets of vectors span the same  $j-1$  dimensional subspace  $L_{j-1}$ . This subspace is distributed according to the Haar measure in the Grassmanian of the  $j-1$  dimensional subspaces of  $\mathbb{R}^n$ .

We consider the column vectors

$$\Delta_j = \sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k = P_{L_{j-1}}(\mathbf{y}_j),$$

where  $P_{L_{j-1}}$  is the orthogonal projection onto  $L_{j-1}$ , and we write

$$\mathbf{y}_j - \sqrt{n}\boldsymbol{\nu}_j = \Delta_j + (\mathbf{y}_j - \Delta_j) - \sqrt{n}\boldsymbol{\nu}_j.$$

Let us call  $\Delta'_j = (\mathbf{y}_j - \Delta_j) - \sqrt{n}\boldsymbol{\nu}_j$  and let us note that  $(\mathbf{y}_j - \Delta_j)$  has the same direction as  $\boldsymbol{\nu}_j$  (by definition of  $\boldsymbol{\nu}_j$ ) so that

$$\Delta'_j = (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\boldsymbol{\nu}_j = (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n})\boldsymbol{\nu}_j,$$

where  $P_{L_{j-1}^\perp}$  is the projection onto the subspace orthogonal to  $L_{j-1}$ . Note that  $\Delta_j$  and  $\Delta'_j$  are orthogonal to each other.

Associated to the  $\Delta_j$ 's and  $\Delta'_j$ 's, for every  $1 \leq i \leq n$  and for every  $1 \leq m \leq n$  we have the (truncated) row vectors

$$G_i^m = (\Delta_j(i))_{j=1}^m = \left( \sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k \rangle \langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle \right)_{j=1}^m = (\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle)_{j=1}^m$$

and

$$\begin{aligned} H_i^m &= (\Delta'_j(i))_{j=1}^m = ((\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle)_{j=1}^m \\ &= \left( (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n})\langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \right)_{j=1}^m. \end{aligned}$$

Then,

$$\|F_i^m\|^2 = \langle G_i^m + H_i^m, G_i^m + H_i^m \rangle = \|G_i^m\|^2 + \|H_i^m\|^2 + 2\langle G_i^m, H_i^m \rangle.$$

We will upper and lower bound  $\|G_i^m\|$  and  $\|H_i^m\|$  outside of a set of exponentially small probability. Moreover, we show that the leading terms of  $\sup_i \|G_i^m\|$  and  $\inf_i \|G_i^m\|$  are equal (and the same happens for  $\|H_i^m\|$ ) and thus the bounds are sharp. After that, we will see that  $\langle G_i^m, H_i^m \rangle$  is negligible when compared with those bounds outside of a set of probability exponentially small. Finally we will get that  $\|F_i^m\|$  is upper and lower bounded by the bounds of  $\sqrt{\|G_i^m\|^2 + \|H_i^m\|^2}$ .

First, we bound  $\|G_i^m\|$ .



**Proposition 3.1.** *With the notation of Theorem 1.1,*

$$\sup_i \|G_i^m\| \leq \sqrt{\frac{\alpha}{2}m} + O(m^\delta)$$

and

$$\inf_i \|G_i^m\| \geq \sqrt{\frac{\alpha}{2}m} - O(m^\delta)$$

with probability exponentially close to 1.

*Proof.* We note that

$$G_{i,j}^m = \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \langle \boldsymbol{\nu}_k, \mathbf{y}_j \rangle = \left\langle \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k, \mathbf{y}_j \right\rangle.$$

Therefore, to obtain the  $j$ -th coordinate of  $G_i^m$  we consider the Haar distributed  $j-1$  dimensional subspace  $L_{j-1} = \text{span}\{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{j-1}\} = \text{span}\{\mathbf{y}_1, \dots, \mathbf{y}_{j-1}\}$ . We project  $\mathbf{e}_i$  onto it and we obtain the vector  $\sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k$ . Independently, we consider a random Gaussian vector  $\mathbf{y}_j$  and calculate the inner product

$$\left\langle \sum_{k=1}^{j-1} \langle \mathbf{e}_i, \boldsymbol{\nu}_k \rangle \boldsymbol{\nu}_k, \mathbf{y}_j \right\rangle = \langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \rangle.$$

The independence of  $\mathbf{y}_j$  with respect to  $\mathbf{y}_1, \dots, \mathbf{y}_{j-1}$  guarantees that

$$G_{i,j}^m = \langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \rangle \text{ is distributed like } \|P_{L_{j-1}}(\mathbf{e}_i)\| g_j = \left( \sum_{k=1}^{j-1} \langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle^2 \right)^{\frac{1}{2}} g_j,$$

where  $g_j$  is a standard normal variable, independent of  $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_{j-1}$ , and, therefore, independent also of all of the previous  $g_{j'}, j' < j$ .

Hence, with the notation  $\langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle = \nu_{k,i}$  we have that  $\|G_i^m\|^2$  has the same distribution as  $\sum_{j=2}^m \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2$ .

The fact that the factors  $\sum_{k=1}^{j-1} \nu_{k,i}^2$  multiplying each of the  $g_j$ 's are not constant and depend on the previous  $g_{j'}$ 's makes it impossible to apply a concentration bound directly. We circumvent this difficulty by grouping the sum in blocks with a constant factor. This increases the total sum by a very small amount.

We partition the set  $\{2, \dots, m\}$  in  $N$  blocks of size  $h = \frac{m-1}{N}$ . Then, we have

$$(3.2) \quad \sum_{j=2}^m \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2 = \sum_{l=1}^N \sum_{j=(l-1)h+2}^{lh+1} \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2 \leq \sum_{l=1}^N \left( \sum_{j=(l-1)h+2}^{lh+1} g_j^2 \right) \left( \sum_{k=1}^{lh} \nu_{k,i}^2 \right).$$

Note that  $(\nu_{k,i})_{k=1}^{lh}$  can be seen as the projection of the vector  $\mathbf{e}_i$  onto a random Haar distributed subspace of dimension  $lh$ . Now we can apply Proposition 2.2, Proposition 2.4 and the union bound, and we get that, for every  $0 < \rho < 1$ ,

$$\mathcal{G}_{n^2} \left\{ \|G_i^m\|^2 \geq \sum_{l=1}^N \left( \frac{1}{(1-\rho)} h \right) \left( \frac{1}{(1-\rho)^2} \frac{lh}{n} \right) \right\} \leq 2Ne^{-\frac{\rho^2 h}{4}}.$$

Then, using the union bound on the  $i$ 's and the definitions of  $\alpha$  and  $N$  we have that, with probability greater than  $1 - 2n\frac{m}{h}e^{-\frac{\rho^2 h}{4}}$ ,

$$(3.3) \quad \sup_i \|G_i^m\|^2 \leq \frac{1}{(1-\rho)^3} \frac{h^2}{n} \frac{N(N+1)}{2} \leq \frac{1}{(1-\rho)^3} \frac{\alpha}{2} (m+h).$$

Different choices of  $h, \rho$  yield now different versions of our result. For instance, we can choose  $h = m^{1/2}$ ,  $\rho = m^{-1/5}$  and we have  $\|G_i^m\|^2 \leq \frac{\alpha}{2}m + O(m^{4/5})$  with probability  $1 - 2n\sqrt{m}e^{-\frac{m^{1/10}}{4}}$ . Easy computations show that  $\|G_i^m\|^2 \leq \frac{\alpha}{2}m + O(m^{2/3+\epsilon})$  with probability tending to 0 exponentially in  $m^\epsilon$ .

We can also choose  $h = \frac{\epsilon}{2}m$  and  $\rho = \frac{\epsilon}{8}$  and, using the Taylor expansion of  $\frac{1}{(1-\rho)^3}$ , we get that

$$\|G_i^m\|^2 \leq (1+\epsilon)\frac{\alpha}{2}m,$$

with probability greater than  $1 - \frac{4n}{\epsilon}e^{-\frac{\epsilon^3 m}{2^9}}$ .

For the sake of clarity, we have written Equation (3.2) as if  $N$  and  $h$  were always integers. If they were not, we can take  $N' = [N] + 1$  and  $h' = [h] + 1$ . This adds at most  $[N] + [h] + 1$  terms to the previous sum. It is very easy to see that this extra terms do not change the above estimates.

This upper bounds  $\|G_i^m\|$ .

Similar reasonings prove the lower bound. To do this, one replaces Equation (3.2) by

$$\sum_{j=2}^m \sum_{k=1}^{j-1} \nu_{k,i}^2 g_j^2 \geq \sum_{l=2}^N \sum_{j=(l-1)h+2}^{lh+1} \sum_{k=1}^{(l-1)h+1} \nu_{k,i}^2 g_j^2,$$

and proceeds similarly as with the upper bound. □

We proceed now to bound  $\|H_i^m\|$ .

**Proposition 3.2.** *With the notation of Theorem 1.1,*

$$\sup_i \|H_i^m\| \leq \sqrt{\left(2 - \alpha/2 - \frac{4}{3} \frac{(1 - (1-\alpha)^{3/2})}{\alpha}\right) m + O(m^\delta)}$$

and

$$\inf_i \|H_i^m\| \geq \sqrt{\left(2 - \alpha/2 - \frac{4}{3} \frac{(1 - (1 - \alpha)^{3/2})}{\alpha}\right) m} - O(m^\delta),$$

with probability exponentially close to 1.

*Proof.* Recall that

$$H_i^m = ((\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})\nu_{j,i})_{j=1}^m,$$

where  $\mathbf{y}_j - \Delta_j$  is the projection of  $\mathbf{y}_j$  onto the  $n - (j - 1)$  dimensional subspace orthogonal to the subspace  $L_{j-1} = \text{span}\{\mathbf{y}_1, \dots, \mathbf{y}_{j-1}\}$ . We will first bound the term  $(\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 = (\|P_{L_{j-1}^\perp} \mathbf{y}_j\| - \sqrt{n})^2$ . For that, we need an auxiliary Lemma which we will also use later.

**Lemma 3.3.** *With the notation above, we have*

(i) *For every  $0 < \rho < 1$  and for every  $m < n$*

$$\left(\sqrt{n} - \|P_{L_{j-1}^\perp} \mathbf{y}_j\|\right)^2 \leq \left(\sqrt{n} - (1 - \rho)^{\frac{1}{2}} \sqrt{n - j + 1}\right)^2 \quad \text{for } 1 \leq j \leq m,$$

*except for a set  $Z_1$  with  $\mathcal{G}_{n^2}(Z_1) \leq m \left(e^{-\frac{\rho^2(n-m+1)}{4}} + e^{-\frac{\rho^2(n-m+1)}{16}}\right)$ .*

(ii) *Let  $0 < \rho_0 < 1$  and  $j_0 \in \mathbb{N}$  such that  $(1 - \rho_0)^{-1}(n - j_0 + 1) \leq n$ . Then, for every  $m < n$*

$$\left(\sqrt{n} - \|P_{L_{j-1}^\perp} \mathbf{y}_j\|\right)^2 \geq \left(\sqrt{n} - (1 - \rho_0)^{-\frac{1}{2}} \sqrt{n - j + 1}\right)^2 \quad \text{for } j_0 < j \leq m,$$

*except for a set  $Z_2$  with  $\mathcal{G}_{n^2}(Z_2) \leq (m - j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}}$ .*

*Proof.* First we prove (i). We choose  $\epsilon = \rho/2$  in Equation (2.3) and  $\epsilon = \rho$  in Equation (2.3) and we get

$$(1 - \rho)^{\frac{1}{2}} \sqrt{n - j + 1} - \sqrt{n} \leq \|P_{L_{j-1}^\perp} \mathbf{y}_j\| - \sqrt{n} \leq \left(1 - \frac{\rho}{2}\right)^{-\frac{1}{2}} \sqrt{n - j + 1} - \sqrt{n}$$

except for a set of measure  $e^{-\frac{\rho^2(n-j+1)}{4}} + e^{-\frac{\rho^2(n-j+1)}{16}}$ . Using the fact that for  $0 < \rho < 1$   $1 - (1 - \rho)^{\frac{1}{2}} \geq (1 - \frac{\rho}{2})^{\frac{-1}{2}} - 1$  we have

$$\left|(1 - \frac{\rho}{2})^{\frac{-1}{2}} \sqrt{n - j + 1} - \sqrt{n}\right| \leq \left|(1 - \rho)^{\frac{1}{2}} \sqrt{n - j + 1} - \sqrt{n}\right|.$$

Then, taking squares and applying a union bound we get (i).

The proof of (ii) follows from Equation (2.3), the extra condition on  $\rho_0$  and  $j_0$  and the union bound.  $\square$

Now, in order to upper bound  $\|H_i^m\|^2$  we first consider the case  $\alpha < 1$ . As in the proof of Theorem 3.1, we partition the set  $\{1, \dots, m\}$  in  $N$  blocks of size  $h = \frac{m}{N}$ . Then, using Lemma 3.3.(i), we write

$$\begin{aligned}
 (3.4) \quad \|H_i^m\|^2 &= \sum_{j=1}^m (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 \\
 &\leq \sum_{l=1}^N \sum_{j=(l-1)h+1}^{lh} \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-lh+1} \right)^2 \nu_{j,i}^2 \\
 &= \sum_{l=1}^N \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-lh} \right)^2 \sum_{j=(l-1)h+1}^{lh} \nu_{j,i}^2,
 \end{aligned}$$

outside of  $Z_1$ .

On the other hand, considering  $(\nu_{k,i})_{i=(l-1)h+1}^{lh}$  as the projection of  $\mathbf{e}_i$  onto a random subspace of dimension  $lh$ , Proposition 2.4 tells us that, for every  $1 \leq i \leq n$  and  $1 \leq l \leq N$ ,

$$\sum_{j=(l-1)h+1}^{lh} \nu_{j,i}^2 \leq \frac{1}{(1-\rho')^2} \frac{h}{n}$$

except for a set  $Z'_1$  with  $\mathcal{G}_{n^2}(Z'_1) \leq nNe^{-\frac{\rho'^2 h}{4}}$ .

So, we have that, outside of  $Z_1 \cup Z'_1$ ,

$$\begin{aligned}
 \|H_i^m\|^2 &\leq \frac{1}{(1-\rho')^2} \frac{h}{n} \sum_{l=1}^N \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-lh} \right)^2 \\
 &= \frac{1}{(1-\rho')^2} \frac{h}{n} \left[ nN + (1-\rho) \left( nN - h \frac{N(N+1)}{2} \right) \right. \\
 &\quad \left. - 2(1-\rho)^{\frac{1}{2}} \sqrt{n} \sum_{l=1}^N \sqrt{n-lh} \right].
 \end{aligned}$$

We can bound

$$\sum_{l=1}^N \sqrt{n-lh} \geq \int_1^N \sqrt{n-xh} dx = \frac{2}{3h} ((n-h)^{3/2} - (n-Nh)^{3/2}).$$

Then, putting this together with the definitions of  $\alpha$  and  $N$ , we get that

$$\begin{aligned}
 \|H_i^m\|^2 &\leq \frac{1}{(1-\rho')^2} m \left[ 1 + (1-\rho) \left( 1 - \frac{\alpha}{2} - \frac{\alpha h}{2m} \right) \right. \\
 &\quad \left. - (1-\rho)^{\frac{1}{2}} \frac{4}{3\alpha} \left( \left( 1 - \frac{\alpha h}{m} \right)^{3/2} - (1-\alpha)^{3/2} \right) \right],
 \end{aligned}$$

with probability greater than

$$(3.5) \quad 1 - \frac{m^2}{\alpha h} e^{-\frac{\rho^2 h}{4}} - m \left( e^{-\frac{\rho^2(n-m)}{4}} + e^{-\frac{\rho^2(n-m)}{16}} \right).$$

Again, different choices of  $h, \rho, \rho'$  yield now different versions of our result. For instance, taking  $h = m^{1/2}$ ,  $\rho = \rho' = m^{-1/5}$  we get

$$\|H_i^m\|^2 \leq \left( 2 - \alpha/2 - \frac{4(1 - (1 - \alpha)^{3/2})}{\alpha} \right) m + O(m^{4/5}),$$

with probability tending to one exponentially in  $m^{1/10}$ . Easy computations also show that  $\|G_i^m\|^2 \leq \frac{\alpha}{2}m + O(m^{2/3+\epsilon})$  with probability tending to 0 exponentially in  $m^\epsilon$ .

The reasonings above do not apply directly to the case  $\alpha = 1$ , as the bound of the probability in Equation (3.5) becomes trivial in that case. To overcome this issue, in case  $\alpha = 1$  we consider  $h = \frac{n - \sqrt{n}}{N}$  and rewrite Equation (3.4) as

$$\begin{aligned} \|H_i^m\|^2 &\leq \sum_{j=1}^n (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 \\ &= \sum_{j=1}^{n-\sqrt{n}} (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 + \sum_{j=n-\sqrt{n}+1}^n (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 \\ &\leq \sum_{l=1}^N \sum_{j=(l-1)h+1}^{lh} \left( \sqrt{n} - (1 - \rho)^{-\frac{1}{2}} \sqrt{n - lh + 1} \right)^2 \nu_{j,i}^2 + \sum_{j=n-\sqrt{n}+1}^n n \nu_{j,i}^2, \end{aligned}$$

outside of the set  $Z_1$  defined in Lemma 3.3 in the case  $m = n - \sqrt{n}$ .

The first summand is treated as previously where now  $m = n - \sqrt{n}$ . We note that, using Proposition 2.4 and the union bound once again, the second summand verifies, with probability greater than  $1 - ne^{-\frac{\rho^2 \sqrt{n}}{4}}$ ,

$$n \sum_{j=n-\sqrt{n}+1}^n \nu_{j,i}^2 \leq n \frac{1}{(1 - \rho^2)\sqrt{n}} = \frac{\sqrt{n}}{(1 - \rho^2)}.$$

This only adds an  $O(\sqrt{n})$  term which does not modify the result. This finishes the proof of the upper bound.

For the lower bound we reason similarly. Consider  $j_0, \rho_0$  as in Lemma 3.3.(ii). Then, with probability  $1 - n(m - j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}}$ , for every  $1 \leq i \leq n$  we have that

$$(3.6) \quad \begin{aligned} \|H_i^m\|^2 &\geq \sum_{j=j_0+1}^m (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 \\ &\geq \sum_{j=j_0+1}^m \left( \sqrt{n} - (1 - \rho_0)^{-1/2} \sqrt{n - j + 1} \right)^2 \nu_{j,i}^2. \end{aligned}$$

Partitioning the set  $\{j_0 + 1, \dots, m\}$  in  $N$  blocks of size  $h = \frac{m-j_0}{N}$  and using similar reasonings to those used for the upper bound in the case  $\alpha < 1$  we obtain

$$\inf_i \|H_i^m\|^2 \geq (1 - \rho')^2(m - j_0) \left[ 1 + (1 - \rho_0) \left( 1 - \frac{\alpha}{2} + \frac{\alpha(j_0 + h)}{2m} \right) - \frac{4(1 - \rho_0)^{\frac{1}{2}}}{3(\alpha - \frac{\alpha j_0}{m})} \left( \left( 1 + \frac{\alpha h}{m} \right)^{3/2} - \left( 1 - \alpha + \frac{\alpha j_0}{m} \right)^{3/2} \right) \right],$$

with probability higher than  $1 - \frac{m}{\alpha}(m - j_0)e^{-\frac{\rho_0^2(n-m+1)}{4}} - \frac{h}{\alpha}e^{-\frac{\rho'^2 h}{4}}$ . As in expressions (3.3) and (3.5) different values of  $j_0, \rho_0, \rho$  and  $h$  give different bounds.

The case  $\alpha = 1$  can be treated as in the upper bound. The terms in (3.6), where  $n - \sqrt{n} + 1 \leq j \leq n$  can only add up to  $O(\sqrt{n})$  and the rest of the terms can be bounded as before.

Note that, as in the proof of Proposition 3.1, we are assuming that  $N, h$  and  $\sqrt{n}$  are integers. If this is not the case we can consider  $N' = [N] + 1$  and  $h' = [h] + 1$  for the upper estimates ( $N' = [N]$  and  $h' = [h]$  for the lower estimates) and  $[\sqrt{n}]$  for the case  $\alpha = 1$ . Adding or subtracting these extra terms will give negligible quantities compared with the sums. This finishes the proof.  $\square$

We now need to prove that  $\langle G_i^m, H_i^m \rangle$  is negligible when compared with  $\|G_i^m\|^2$  and  $\|H_i^m\|^2$ . More precisely, we will use similar techniques to show that  $\langle G_i^m, H_i^m \rangle$  is, with probability exponentially close to 1, of smaller order in  $m$  than  $\|G_i^m\|^2$  and  $\|H_i^m\|^2$ . As shown in Proposition 3.1 and Proposition 3.2 above, each of them is  $\Theta(m)$ . That is, there exist constants  $k_1, k_2$  and  $m_0$  such that for all  $m > m_0$ , we have that  $k_1 m \leq \|G_i^m\|^2 \leq k_2 m$  and analogously for  $\|H_i^m\|^2$ .

**Proposition 3.4.** *With the notation of Theorem 1.1, given  $\epsilon > 0$  we have*

$$\langle G_i^m, H_i^m \rangle = O(m^{\frac{1}{2} + \epsilon}),$$

*with probability exponentially close to 1.*

This proposition, together with Propositions 3.1 and 3.2, finishes the proof of Theorem 1.1.

For the sake of clarity we will first show the following technical lemma that will be used in the proof of Proposition 3.4.

**Lemma 3.5.** *Let  $\mathbf{y}_j$  be a Gaussian vector and  $L_{j-1}$  a Haar distributed subspace of dimension  $j-1$ , then*

$$\Pr \left( \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle^2 > \frac{j-1}{n} m^\epsilon \right) \leq 2e^{-\frac{m^{\epsilon/2}-2}{4}}.$$

*Proof.* First of all note that  $\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle = \langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \rangle$ . We have already shown in Equation (3.1) that  $\langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \rangle^2$  has the same distribution as the term  $\|P_{L_{j-1}}(\mathbf{e}_i)\|^2 g_j^2$ , where  $g_j$  is a standard normal variable. Putting together Lemma 2.1, Proposition 2.4 and the union bound, we get

$$\Pr \left( \langle P_{L_{j-1}}(\mathbf{e}_i), \mathbf{y}_j \rangle^2 > \frac{j-1}{n} m^\epsilon \right) \leq e^{-\frac{j-1}{4}(m^{\epsilon/2}-2)} + e^{-\frac{m^{\epsilon/2}}{2}} \leq 2e^{-\frac{m^{\epsilon/2}-2}{4}}.$$

□

We will also need Hoeffding's inequality [13].

**Proposition 3.6** (Hoeffding's inequality). *Let  $(X_i)_{i=1}^n$  be a family of independent random variables such that  $a_i \leq X_i \leq b_i$  for  $i = 1, \dots, n$ . Let  $S = \sum_{i=1}^n X_i$ . Then, for every  $a > 0$ ,*

$$\Pr(|S - \mathbb{E}(S)| > a) \leq 2e^{-\frac{2a^2}{\sum_i (b_i - a_i)^2}}.$$

*Proof of Proposition 3.4.* Recall that we have

$$G_i^m = \left( \sum_{k=1}^{j-1} \langle \mathbf{y}_j, \boldsymbol{\nu}_k \rangle \langle \boldsymbol{\nu}_k, \mathbf{e}_i \rangle \right)_{j=1}^m = \left( \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle \right)_{j=1}^m$$

and

$$H_i^m = \left( (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \right)_{j=1}^m = \left( (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \right)_{j=1}^m.$$

Therefore,

(3.7)

$$\begin{aligned} \langle G_i^m, H_i^m \rangle &= \sum_{j=1}^m \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \\ &= \sum_{j=1}^m |\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle| (\|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n}) \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle \text{sign}(\langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle). \end{aligned}$$

We claim that the probability distribution of the previous expression is the same as the probability distribution of

$$(3.8) \quad \sum_{j=1}^m |\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle| (\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n}) \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle \epsilon_j,$$

where  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_n$  are independent Gaussian vectors,  $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n$  are the corresponding vectors obtained from the Gram-Schmidt orthonormalization procedure and  $\epsilon_1, \dots, \epsilon_n$  are independent and identically distributed Bernoulli variables taking values in an independent probability space.

In order to see this, let us consider the space  $\mathbb{R}^n \times \dots \times \mathbb{R}^n$  with the Gaussian measure  $\mathcal{G}_n$  on each  $\mathbb{R}^n$ . For each  $j$  we denote by  $\mathbf{z}_j = (z_{k,j})_{k=1}^n$  the Gaussian vector in the corresponding copy of  $\mathbb{R}^n$ . For each  $j \geq 2$  we consider in  $\mathbb{R}^n$  the equivalence relation  $\mathbf{z} \sim_j \mathbf{z}'$  if and only if  $(z_1, \dots, z_{j-1}) = \pm(z'_1, \dots, z'_{j-1})$  and  $(z_j, \dots, z_n) = (z'_j, \dots, z'_n)$ . Then,  $\mathbb{R}^n = (\mathbb{R}^n / \sim_j) \times \{-1, 1\}$ , with the identification  $\mathbf{z}_j = ([\mathbf{z}_j], \sigma_j)$ , with  $\sigma_j \in \{-1, 1\}$ . We define the probability measure  $\mathcal{G}'_n$  on  $\mathbb{R}^n / \sim_j$  by the density  $f'([\mathbf{z}_j]) = 2f(\mathbf{z}_j)$ , where  $f$  is the density of  $\mathcal{G}_n$ , and we call  $\mu$  the uniform probability on  $\{-1, 1\}$ . We clearly have  $\mathcal{G}_n = \mathcal{G}'_n \otimes \mu$ .

Let us now consider a family of independent Gaussian vectors  $(\mathbf{z}_1, \dots, \mathbf{z}_n)$  in the previous probability space. For  $j = 1$  there is no equivalence relation, and we define  $\tilde{\mathbf{y}}_1 = \mathbf{z}_1$ , which is clearly a Gaussian vector. Consequently, we define  $\tilde{\mathbf{v}}_1 = \frac{1}{\|\tilde{\mathbf{y}}_1\|} \tilde{\mathbf{y}}_1$ .

Now, for each  $2 \leq j \leq m$ , we consider  $\tilde{L}_{j-1}$ , the random  $(j-1)$ -dimensional subspace spanned by  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_{j-1}$ . The vectors  $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_{j-1}$  form an orthonormal basis of  $\tilde{L}_{j-1}$ . Hence, we can complete this set to obtain a basis of  $\mathbb{R}^n$ ,  $\{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_{j-1}, \boldsymbol{\nu}_j^*, \dots, \boldsymbol{\nu}_n^*\}$ . The added vectors needed to complete the orthonormal basis can be chosen at will. In general they will change as  $j$  changes.

Let us denote by  $U_j$  the orthogonal matrix whose columns are the vectors of the previous basis. Then, given  $\mathbf{z}_j$ , we define  $\tilde{\mathbf{y}}_j = U_j \mathbf{z}_j$ . Since the orthogonal matrix  $U_j$  is independent of the Gaussian vector  $\mathbf{z}_j$ , we immediately deduce that  $\tilde{\mathbf{y}}_j$  is a Gaussian vector independent of  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_{j-1}$ .

From the Gram-Schmidt orthonormalization procedure we have that

$$\tilde{\mathbf{v}}_j = \frac{\tilde{\mathbf{y}}_j - \sum_{k=1}^{j-1} \langle \tilde{\mathbf{y}}_j, \tilde{\mathbf{v}}_k \rangle \tilde{\mathbf{v}}_k}{\|\tilde{\mathbf{y}}_j - \sum_{k=1}^{j-1} \langle \tilde{\mathbf{y}}_j, \tilde{\mathbf{v}}_k \rangle \tilde{\mathbf{v}}_k\|}.$$

It follows immediately that

$$P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j) = \sum_{k=1}^{j-1} z_{k,j} \tilde{\mathbf{v}}_k \quad \text{and} \quad P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j) = \sum_{k=j}^n z_{k,j} \boldsymbol{\nu}_k^*,$$



where we recall that  $\mathbf{z}_j = (z_{k,j})_{k=1}^n$ .

With the identification  $\mathbf{z}_j = ([\mathbf{z}_j], \sigma_j)$ , it is easy to see that  $|\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle|$  and  $(\|P_{\tilde{L}_{j-1}}^\perp(\tilde{\mathbf{y}}_j)\| - \sqrt{n}) \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle$  do not depend on  $\sigma_2, \dots, \sigma_n$  (or, equivalently, they only depend on the variables  $[\mathbf{z}_j]$ ). Indeed, to see this we note first that it follows from the definitions that the vectors  $\tilde{\mathbf{v}}_j$  are independent of  $\sigma_2, \dots, \sigma_n$ . Next, we note that the dependence of  $\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle$  with respect to  $\sigma_2, \dots, \sigma_n$  is cancelled out by the absolute value.

Hence, expression (3.7) applied to the independent Gaussian vectors  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_n$  has the form

$$\sum_{j=1}^m |\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle| (\|P_{\tilde{L}_{j-1}}^\perp(\tilde{\mathbf{y}}_j)\| - \sqrt{n}) \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle \epsilon_j,$$

where  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_n$  are Gaussian vectors independent of  $\sigma_2, \dots, \sigma_n$  and  $\epsilon_j = \epsilon_j(\sigma_j)$  are independent identically distributed Bernoulli variables.

Equation (3.8) will allow us to apply Proposition 3.6: For fixed  $(\mathbf{z}_1, [\mathbf{z}_2], \dots, [\mathbf{z}_n])$  we can consider the independent random variables (function of  $(\sigma_2, \dots, \sigma_n)$ )

$$\left( |\langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle| (\|P_{\tilde{L}_{j-1}}^\perp(\tilde{\mathbf{y}}_j)\| - \sqrt{n}) \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle \epsilon_j(\sigma_j) \right)_{j=1}^m,$$

Then, Proposition 3.6 gives us that, for fixed  $(\mathbf{z}_1, [\mathbf{z}_2], \dots, [\mathbf{z}_n])$ .

$$(3.9) \quad \mu^{\otimes m} \left( |\langle G_i^m, H_i^m \rangle| \geq a \right) \leq 2e^{-\frac{a^2}{2 \sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle^2 (\|P_{\tilde{L}_{j-1}}^\perp(\tilde{\mathbf{y}}_j)\| - \sqrt{n})^2 \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle^2}}.$$

We consider first the case  $\alpha < 1$ .

It follows from Lemma 3.5, Proposition 2.4, Lemma 3.3 and a union bound argument that

$$(3.10) \quad \begin{aligned} & 2 \sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle^2 (\|P_{\tilde{L}_{j-1}}^\perp(\tilde{\mathbf{y}}_j)\| - \sqrt{n})^2 \langle \tilde{\mathbf{v}}_j, \mathbf{e}_i \rangle^2 \leq \\ & \leq 2 \sum_{j=1}^m m^\epsilon \frac{j-1}{n} (\sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-j+1})^2 \frac{m^\epsilon}{n} \leq m^{2\epsilon+1} \end{aligned}$$

with probability larger than

$$1 - C(m, \rho, \epsilon) := 1 - 2me^{-(\frac{m^{\frac{\epsilon}{2}} - 2}{4})} + m \left( e^{-\frac{\rho^2((1/\alpha-1)m+1)}{4}} + e^{-\frac{\rho^2((1/\alpha-1)m+1)}{16}} \right) + \sqrt{e} m^{\frac{1}{4}m^{2\epsilon}}.$$

(The second inequality in the second line of (3.10) follows from easy calculations).

We note now that  $Pr\left(|\langle G_i^m, H_i^m \rangle| \geq a\right)$  is upper bounded by

$$\begin{aligned} Pr\left(|\langle G_i^m, H_i^m \rangle| \geq a \left| 2 \sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle^2 (\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n})^2 \langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i \rangle^2 \leq m^{2\epsilon+1} \right) \right. \\ \left. + Pr\left(2 \sum_{j=1}^m \langle P_{\tilde{L}_{j-1}}(\tilde{\mathbf{y}}_j), \mathbf{e}_i \rangle^2 (\|P_{\tilde{L}_{j-1}^\perp}(\tilde{\mathbf{y}}_j)\| - \sqrt{n})^2 \langle \tilde{\boldsymbol{\nu}}_j, \mathbf{e}_i \rangle^2 > m^{2\epsilon+1} \right), \right. \end{aligned}$$

where we denote by  $Pr(A|B)$  the probability of the event  $A$  conditioned to  $B$ .

We pick  $\epsilon' > \epsilon$  and we fix  $a = m^{\frac{1}{2}+\epsilon'}$ . Then, Equations (3.9) and (3.10) imply that

$$Pr\left(|\langle G_i^m, H_i^m \rangle| \geq m^{\frac{1}{2}+\epsilon'}\right) \leq 2e^{-\frac{m^{1+2\epsilon'}}{m^{2\epsilon+1}}} + C(m, \rho, \epsilon) = 2e^{m^{-2(\epsilon'-\epsilon)}} + C(m, \rho, \epsilon),$$

which tends exponentially fast to zero as  $n$  grows to infinity.

The case  $\alpha = 1$  has to be considered separately as the bound in the concentration of Lemma 3.3 becomes trivial. In order to overcome this issue we reason as in the proof of Proposition 3.2. We can divide the sum in (3.10) in two terms

$$\begin{aligned} & 2 \sum_{j=1}^{n-\sqrt{n}} \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle^2 \left( \|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n} \right)^2 \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle^2 \\ & + 2 \sum_{j=n-\sqrt{n}+1}^n \langle P_{L_{j-1}}(\mathbf{y}_j), \mathbf{e}_i \rangle^2 \left( \|P_{L_{j-1}^\perp}(\mathbf{y}_j)\| - \sqrt{n} \right)^2 \langle \boldsymbol{\nu}_j, \mathbf{e}_i \rangle^2, \end{aligned}$$

where the first summand is treated as previously giving an upper bound of  $(n - \sqrt{n})^{2\epsilon+1}$  and the second is  $O(n^{1/2+2\epsilon})$ , which is negligible compared with the first. Proceeding as in the case  $\alpha < 1$  the result follows.  $\square$

We prove now Corollary 1.2. Before writing the actual proof, we note that considering the Taylor expansion of the expression in Theorem 1.1 we obtain

$$2 - \frac{4(1 - (1 - \alpha)^{3/2})}{\alpha} = \frac{\alpha}{2} + \frac{\alpha^2}{12} + O(\alpha^3),$$

so that for  $\alpha$  close enough to 0 we would expect

$$\|F_i^m\| \approx \sqrt{\left(2 - \frac{4(1 - (1 - \alpha)^{3/2})}{\alpha}\right)m} \approx \sqrt{\frac{\alpha m}{2}} = \sqrt{\frac{m^2}{2n}},$$

as it is indeed the case.

**Proof of Corollary 1.2:** As in the proof of Theorem 1.1 we consider the decomposition  $F_i^m = G_i^m + H_i^m$  and we treat separately the norms of  $G_i^m$  and  $H_i^m$ .

We start with  $\|G_i^m\|$ . Recall that Equation (3.3) says that, with probability greater than  $1 - 2n \frac{m}{h} e^{-\frac{\rho^2 h}{4}}$ ,

$$\sup_i \|G_i^m\|^2 \leq \frac{1}{(1-\rho)^3} \frac{m}{2n} (m+h).$$

If we make, for instance, the choices  $\rho = m^{-\frac{1}{6}}$ ,  $h = m^{\frac{1}{2}}$ , and take into account that  $\frac{1}{(1-\rho)^3} \leq 1 + 4\rho$  for  $\rho$  small enough, we get that

$$\sup_i \|G_i^m\|^2 \leq \frac{m^2}{2n} + O\left(\frac{m^{\frac{11}{6}}}{n}\right),$$

with probability greater than  $1 - 2nm^{\frac{1}{2}} e^{-\frac{1}{4}}$ .

Similar reasonings show that

$$\inf_i \|G_i^m\|^2 \geq \frac{m^2}{2n} - O\left(\frac{m^{\frac{11}{6}}}{n}\right),$$

with probability greater than  $1 - 2nm^{\frac{1}{2}} e^{-\frac{1}{4}}$ .

Now we bound the norm of  $H_i^m$ . It follows from the proof of Proposition 3.2 (see Equation (3.4) and Lemma 3.3) that

$$\begin{aligned} \|H_i^m\|^2 &= \sum_{j=1}^m (\|\mathbf{y}_j - \Delta_j\| - \sqrt{n})^2 \nu_{j,i}^2 \\ &\leq \sum_{j=1}^m \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-j+1} \right)^2 \nu_{j,i}^2 \end{aligned}$$

holds for every  $i = 1, \dots, n$ , outside of a set  $Z_1$  with  $\mathcal{G}_{n^2}(Z_1) \leq nm \left( e^{-\frac{\rho^2(n-m+1)}{4}} + e^{-\frac{\rho^2(n-m+1)}{16}} \right)$ , where we have applied a union bound argument in  $i$ .

For every  $1 \leq j \leq m$ , we have

$$\left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-j+1} \right)^2 \leq \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-m} \right)^2.$$

Using this and Proposition 2.4 as in the proof of Proposition 3.2, we get that

$$\begin{aligned} \|H_i^m\|^2 &\leq \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-m} \right)^2 \sum_{j=1}^m \nu_{j,i}^2 \\ &\leq \frac{1}{(1-\rho')^2} \frac{m}{n} \left( \sqrt{n} - (1-\rho)^{\frac{1}{2}} \sqrt{n-m} \right)^2, \end{aligned}$$

outside of the sets  $Z_1$  and  $Z'_1$ , where  $Z_1$  is as above and  $\mathcal{G}_{n^2}(Z'_1) \leq ne^{-\frac{\rho'^2 m}{4}}$ .

We multiply and divide by the conjugate and lower bound the denominator by  $n$ , and we get

$$\sup_i \|H_i^m\|^2 \leq \frac{1}{(1-\rho')^2} \frac{m^2}{n} \left( \rho^2 \frac{n}{m} + 2\rho + \frac{m}{n} \right).$$

If we choose now, for instance,  $\rho = \frac{m^{\frac{1}{3}}}{n^{\frac{1}{2}}}$ ,  $\rho' = m^{-\frac{1}{3}}$  and using that  $\frac{m}{n} \rightarrow 0$ , we have that

$$\sup_i \|H_i^m\|^2 = o\left(\frac{m^2}{n}\right)$$

with probability exponentially small in  $m$ . Thus, in this regime  $\|H_i^m\|$  is negligible compared to  $\|G_i^m\|$ .

Since

$$\|G_i^m\| - \|H_i^m\| \leq \|F_i^m\| \leq \|G_i^m\| + \|H_i^m\|,$$

this finishes the proof.  $\square$

#### 4. THE SUPREMUM NORM

In this section we use Theorem 1.1 to prove Theorem 1.3, that describes the asymptotic probabilistic behaviour of  $\epsilon_n(m) = \sup_{1 \leq i \leq n, 1 \leq j \leq m} |y_{i,j} - \sqrt{n}u_{i,j}|$ .

If we choose  $m_n = \frac{\beta n}{\log n}$  or  $m_n = o\left(\frac{n}{\log n}\right)$  we immediately obtain Corollary 4.2. This result should be compared with [14, Theorem 3]: In it, Jiang showed that if  $Y_n$  is a Gaussian random matrix and  $U_n$  is the result of its Gram-Schmidt orthonormalization, then  $\epsilon_n(m)$  converges to 0 in probability if and only if  $m_n = o\left(\frac{n}{\log n}\right)$ , and he also showed that if  $m_n = \frac{\beta n}{\log n}$  then  $\epsilon_n(m)$  converges in probability to  $2\sqrt{\beta}$ . Our Corollary 4.2 shows the existence of couplings between a Gaussian matrix  $Y_n$  and a Haar distributed orthogonal matrix  $U_n$  such that  $\epsilon_n(m)$  also converges to 0 in probability if and only if  $m_n = o\left(\frac{n}{\log n}\right)$  but now, when  $m_n = \frac{\beta n}{\log n}$ , the upper bound for  $\epsilon_n(m)$  converges in probability to  $\sqrt{2\beta}$ .

Before we start our reasonings, we state and prove for completeness a lemma which is well known, but for which we have not found an explicit reference.

**Lemma 4.1.** *Let  $\{\mathbf{w}_j = (w_{ij})_{i=1}^n\}_{j=1}^m$  be  $m$  unitary vectors each of which is uniformly distributed in the sphere of  $\mathbb{R}^n$ . Then, for any  $\epsilon > 0$  we have*

$$Pr \left( \sup_{i,j} |w_{i,j}| > (1+\epsilon) \frac{\sqrt{2 \log(nm)}}{\sqrt{n}} \right) \xrightarrow{n} 0,$$

independently of the number of vectors  $m$ . On the other hand, if  $m \leq \alpha n$  for some fixed number  $\alpha$  we have

$$Pr \left( \sup_{i=1, \dots, n} |w_{i,j}| \geq (1 - \epsilon) \frac{\sqrt{2 \log n}}{\sqrt{n}} \text{ for all } j \in \{1, \dots, m\} \right) \xrightarrow{n} 1.$$

*Proof.* In order to prove the first expression, we consider the function that projects a unitary vector in  $\mathbb{R}^n$  onto its  $i$ th coordinate. This function has Lipschitz constant 1 and its median is 0. Thus, a straightforward consequence of Levy's lemma [17] shows that, for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ ,

$$Pr(|w_{i,j}| > t) \leq \sqrt{\frac{\pi}{2}} e^{-(n-1)t^2/2}.$$

Taking  $t = (1 + \epsilon) \frac{\sqrt{2 \log(nm)}}{\sqrt{n}}$  and applying a union bound we get

$$Pr \left( \sup_{i,j} |w_{i,j}| > (1 + \epsilon) \frac{\sqrt{2 \log(nm)}}{\sqrt{n}} \right) \leq \sqrt{\frac{\pi}{2}} (nm)^{-2\epsilon - \epsilon^2 + \frac{(1+\epsilon)^2}{n}} \xrightarrow{n} 0.$$

For the second part of the statement we first consider an arbitrary Gaussian vector  $\mathbf{g} = (g_1, \dots, g_n) \in \mathbb{R}^n$ . It follows from the independence of the coordinates of  $\mathbf{g}$  and the lower bound in Lemma 2.1 that, for any  $t > 0$ ,

$$(4.1) \quad Pr \left( \sup_{i=1, \dots, n} |g_i| < t \right) = (Pr(|g_1| < t))^n \leq \left( 1 - \frac{2}{\sqrt{2\pi}} \frac{t}{1+t^2} e^{-t^2/2} \right)^n.$$

Using the fact that a uniform unitary vector distributes like a normalized Gaussian vector together with a union bound, we deduce that for every  $t, s > 0$  we have

$$\begin{aligned} & Pr \left( \exists j \in \{1, \dots, m\} : \sup_i |w_{i,j}| < t \right) \\ & \leq m Pr \left( \sup_i |w_{i,1}| < t \right) = m Pr \left( \frac{\sup_i |g_i|}{\|\mathbf{g}\|} < t \right) \\ & \leq m Pr \left( \left\{ \sup_i |g_i| < ts \right\} \cap \left\{ \|\mathbf{g}\| \leq s \right\} \right) + m Pr(\|\mathbf{g}\| > s) \\ & \leq m Pr \left( \sup_i |g_i| < ts \right) + m Pr(\|\mathbf{g}\| > s). \end{aligned}$$

We fix  $t = (1 - \epsilon) \frac{\sqrt{2 \log n}}{\sqrt{n}}$ ,  $s = \frac{\sqrt{n}}{\sqrt{1-\delta}}$  and we apply Proposition 2.2 to get that, for  $0 < \delta < 1$ ,  $0 < \epsilon < 1$ , with  $\frac{(1-\epsilon)}{\sqrt{1-\delta}} < 1$ ,

$$\begin{aligned} & Pr \left( \exists j \in \{1, \dots, m\} : \sup_i |w_{i,j}| < (1 - \epsilon) \frac{\sqrt{2 \log n}}{\sqrt{n}} \right) \\ & \leq m Pr \left( \sup_i |g_i| < \frac{(1 - \epsilon)}{\sqrt{1 - \delta}} \sqrt{2 \log n} \right) + m e^{-\delta^2 n/4}. \end{aligned}$$

Using Equation (4.1) we easily conclude that

$$Pr \left( \exists j \in \{1, \dots, m\} : \sup_{i=1, \dots, n} |w_{i,j}| < (1 - \rho) \frac{\sqrt{2 \log n}}{\sqrt{n}} \right) \xrightarrow{n} 0,$$

provided that  $m$  grows as in the statement. Considering the complementary event we obtain the desired result.  $\square$

Theorem 1.1 gives us control over the Euclidean norm of the rows  $F_i^m$  of  $Y_n - \sqrt{n}U_n$ . We will use these estimates to obtain information about the supremum

$$\epsilon_n(m) = \sup_{1 \leq i \leq n, 1 \leq j \leq m} |y_{i,j} - \sqrt{n}u_{i,j}|.$$

First of all we note that  $U_n$  is the Gram-Schmidt orthonormalization of  $Y_n$ . Therefore, the columns of  $Y_n - \sqrt{n}U_n$  are not equally distributed. For instance, it is very easy to see that, with very high probability, their Euclidean norm is strictly increasing. In turn, this implies that the coordinates of each of the  $F_i^m$  are not equally distributed. To avoid this problem, we will randomly choose a slightly better coupling than the one given by the Gram-Schmidt orthonormalization procedure.

**Proof of Theorem 1.3:** Let  $Y_n, U_n$  be as in Theorem 1.1. We consider a Haar distributed orthogonal random matrix  $V_m \in \mathcal{O}(m)$ . We define the orthogonal matrix  $V = (v_{i,j})_{i,j=1}^n \in \mathcal{O}(n)$  by

$$V = \begin{pmatrix} V_m & 0 \\ 0 & I_{n-m} \end{pmatrix}.$$

We now define  $Y' = YV$ ,  $U' = UV$ . Due to the orthogonal invariance of both the Gaussian distribution and the Haar distribution, we have that  $Y'$  is a random Gaussian matrix and  $U'$  is Haar distributed in the orthogonal group. Note that  $U'$  is *not* the Gram-Schmidt orthonormalization of  $Y'$ .

We have now that  $Y'_n - \sqrt{n}U'_n = (Y_n - \sqrt{n}U_n)V$ . Call  $F_{i,j}$  to the  $j^{th}$  coordinate of the vector  $F_i^m$  defined as in Theorem 1.1. Then the first  $m$  coordinates of the

$i^{th}$  row of  $Y'_n - \sqrt{n}U'_n$  form the vector  $\mathbf{x}_i = (x_{i,j})_{j=1}^m$ , where

$$x_{i,j} = \sum_{k=1}^m F_{i,k} v_{k,j}.$$

Therefore  $\mathbf{x}_i \in \mathbb{R}^m$  is a vector whose direction is uniformly distributed and it verifies  $\|\mathbf{x}_i\| = \|F_i^m\|$ . That is, for every  $1 \leq i \leq n$ ,  $\frac{\mathbf{x}_i}{\|F_i^m\|}$  is a unitary vector uniformly distributed.

We will first prove the upper bound of  $\epsilon_n(m)$ . It follows from the first part of Lemma 4.1 that, for every  $t > 0$ ,

$$Pr \left( \sup_{i,j} \frac{|x_{i,j}|}{\|F_i^m\|} > (1 + \varepsilon) \frac{\sqrt{2 \log(nm)}}{\sqrt{m}} \right) \xrightarrow{n} 0.$$

We have that

$$(4.2) \quad Pr \left( \epsilon_n(m) > (1 + \varepsilon) \sup_i \|F_i^m\| \frac{\sqrt{2 \log(nm)}}{\sqrt{m}} \right) \xrightarrow{n} 0.$$

We recall that, according to Theorem 1.1, there exists  $0 < \delta < \frac{1}{2}$  such that

$$\sup_i \|F_i^m\| \leq \sqrt{\left(2 - \frac{4(1 - (1 - \alpha)^{3/2})}{\alpha}\right) m} + O(m^{\frac{1}{2} - \delta}) := \sqrt{\varphi(\alpha)m} + O(m^{\frac{1}{2} - \delta})$$

with probability exponentially close to 1. Putting this together with Equation (4.2) we get the upper bound.

For the lower bound, we will consider the column vectors  $\tilde{\mathbf{x}}_j = (x_{i,j})_{i=1}^n$ , with  $j = 1, \dots, m$ , corresponding to the matrix  $Y'_n - \sqrt{n}U'_n$ . We first note that the normalized vectors  $\frac{\tilde{\mathbf{x}}_j}{\|\tilde{\mathbf{x}}_j\|}$  can be assumed to be uniformly distributed for every  $j$ . Indeed, this follows from the fact that the initial coupling defined via the Schmidt orthonormalization procedure is invariant under left multiplication by an orthogonal matrix. Since left and right multiplication are commutative operations, the invariance under left multiplication by an orthogonal matrix also holds for the random matrix  $Y'_n - \sqrt{n}U'_n$ . Then, according to Lemma 4.1 we have

$$Pr \left( \sup_{i=1, \dots, n} |x_{i,j}| \geq (1 - \epsilon) \|\tilde{\mathbf{x}}_j\| \frac{\sqrt{2 \log n}}{\sqrt{n}} \text{ for all } j \in \{1, \dots, m\} \right) \xrightarrow{n} 1.$$

Now, we will use the trivial fact that  $\sum_{j=1}^m \|\tilde{\mathbf{x}}_j\|^2 = \sum_{i=1}^n \|\mathbf{x}_i\|^2 = \sum_{i=1}^n \|F_i^m\|^2$  and the lower bound provided in Theorem 1.1 to conclude that  $\sum_{j=1}^m \|\tilde{\mathbf{x}}_j\|^2 \geq n(\sqrt{\varphi(\alpha)m} - O(m^\delta))^2$ , so

$$\sup_{j=1, \dots, m} \|\tilde{\mathbf{x}}_j\| \geq \sqrt{n}(\sqrt{\varphi(\alpha)} - O(m^{\delta - \frac{1}{2}})),$$

with probability exponentially close to 1 as  $n$  grows to infinity. Putting the two previous estimates together we easily deduce that

$$\Pr \left( \epsilon_n(m) \geq (1 - \varepsilon) \left( \sqrt{\varphi(\alpha)} - O(m^{\delta - \frac{1}{2}}) \right) \sqrt{2 \log n} \right) \xrightarrow{n} 1,$$

which finishes the proof.  $\square$

**Remark 4.1.** We expect the lower bound of Theorem 1.3 to be  $\sqrt{\varphi(\alpha)} \sqrt{2 \log(nm)}$ . To prove that, one needs to overcome the lack of independence of the rows  $F_i^m$ .

In our previous results,  $\alpha = \frac{m}{n}$  was a constant number. One can easily check that the proof of Theorem 1.3 works in a more general context. In particular, the same argument works if one considers  $m_n = \lfloor \beta \frac{n}{\ln n} \rfloor$  for any constant  $\beta > 0$ . Note that in this case  $\alpha_n = \frac{\beta}{\ln n} \xrightarrow{n} 0$ . With this at hand, we can state and prove the announced improvement of [14, Theorem 3].

**Corollary 4.2.** *For each  $n \geq 2$ , there exist matrices  $Y'_n = (y'_{ij})_{i,j=1}^n$  and  $U'_n = (u'_{i,j})_{i,j=1}^n$  whose  $2n^2$  entries are real random variables defined on the same probability space such that*

- (i) *the law of  $U'_n$  is the normalized Haar measure on the orthogonal group  $\mathcal{O}(n)$ ;*
- (ii)  *$\{y'_{i,j}; 1 \leq i, j \leq n\}$  are independent standard normals;*
- (iii) *set*

$$\epsilon_n(m) = \max_{1 \leq i \leq n, 1 \leq j \leq m} |\sqrt{n} u'_{i,j} - y'_{i,j}|$$

*for  $m = 1, 2, \dots, n$ . Then, for any constant  $\beta > 0$  and  $m_n = \lfloor \frac{\beta n}{\log n} \rfloor$ , for any  $\varepsilon > 0$ ,  $\epsilon_n(m)$  belongs to the interval*

$$(\sqrt{\beta} - \varepsilon, \sqrt{2\beta} + \varepsilon)$$

*with probability  $1 - o(1)$ . Moreover, if we make  $m_n = o(\frac{n}{\ln n})$ , then we have that  $\epsilon_n(m) \rightarrow 0$  in probability as  $n \rightarrow \infty$ .*

*Proof.* Let  $Y'_n, U'_n$  be as in Theorem 1.3. As previously discussed, Theorem 1.3 also holds for  $m_n = \lfloor \frac{\beta n}{\ln n} \rfloor$  with  $\beta > 0$  and in this case  $\alpha_n = \frac{\beta}{\ln n}$ . Then, there exists  $0 < \delta < \frac{1}{2}$  such that for any  $\varepsilon', \varepsilon'' > 0$  and for  $n$  large enough we have

$$\epsilon_n(m) \leq (1 + \varepsilon') \left( \sqrt{\frac{\alpha_n}{2} + \alpha_n^2 + O(m_n^{-\delta})} \right) \sqrt{2 \log(nm_n)} \leq (1 + \varepsilon') (\sqrt{2\beta} + \varepsilon'')$$

with probability tending to 1. Here, we have used that for any  $0 < \alpha \leq 1$  the function  $\varphi(\alpha) = \left( 2 - \frac{4}{3} \frac{(1 - (1 - \alpha)^{3/2})}{\alpha} \right)$  is upper bounded by  $\frac{\alpha}{2} + \alpha^2$ . This proves the upper bound in the first part of the  $\lfloor \frac{\beta n}{\log n} \rfloor$  statement. For the lower bound we reason similarly using the lower bound in Theorem 1.3. The statement about



$m_n = o(\frac{n}{\ln n})$  follows straightforward by considering a sequence  $b_n \rightarrow 0$  and the previous estimate.  $\square$

### ACKNOWLEDGMENTS

We would like to thank the referee for many useful comments which helped to improve the readability of this paper and for suggesting us the explicit inclusion of Corollary 1.2.

Author's research was supported by Spanish research projects MINECO (grants MTM2011-26912 and MTM2014-54240-P), Comunidad de Madrid (grant QUITEMAD+-CM, ref. S2013/ICE-2801), "Ramón y Cajal" program and ICMAT Severo Ochoa project SEV-2015-0554 (MINECO), and the European CHIST-ERA project CQC (funded partially by MINECO grant PRI-PIMCHI-2011-1071).

### REFERENCES

- [1] A. Ambainis, A. Backurs, K. Balodis, D. Kravcenko, R. Ozols, J. Smotrovs, M. Virza, *Quantum strategies are better than classical in almost any XOR game*, Automata, Languages, and Programming Lecture Notes in Computer Science Volume 7391, 2012, 25-37. arXiv:1112.3330.
- [2] E. Borel, *Introduction géométrique à quelques théories physiques*, Gauthier-Villars, Paris, 1906.
- [3] B. Collins, *Intégrales matricielles et probabilités non-commutatives*, Thèse de doctorat, Université Pierre et Marie Curie-Paris VI (2003).
- [4] B. Collins, M. Stolz, *Borel theorems for random matrices from the classical compact symmetric spaces*, Ann. Probab. 36(3), 876-895 (2008).
- [5] A. D'Aristotile, P. Diaconis, C. M. Newman *Brownian motion and the classical groups*, Lecture Notes-Monograph Series, 97-116 (2003).
- [6] S. Dasgupta, A. Gupta, *An elementary proof of a theorem of Johnson and Lindenstrauss*, Random Struct. Alg. 22(1), 60-65 (2003).
- [7] A. Defant and K. Floret, *Tensor Norms and Operator Ideals*, North-Holland, (1993).
- [8] P. W. Diaconis, *Patterns in eigenvalues: The 70th Josiah Willard Gibbs Lecture*, Bull. Amer. Math. Soc. 40, 155-178 (2003).
- [9] P. W. Diaconis, M. L. Eaton, S. L. Lauritzen, *Finite de Finetti theorems in linear models and multivariate analysis*, Scand. J. Statist. 19 289-315 (1992).
- [10] P. W. Diaconis, D. Freedman, *A dozen de Finetti-style results in search of a theory*, Ann. Inst. H. Poincaré Probab. Statist. 23 397-423 (1987).
- [11] L. Gallardo, *Au sujet du contenu probabiliste d'un lemme d'Henri Poincaré*, Ann. Univ. Clermont 69 192-197 (1983).
- [12] C. E. González-Guillén, C. H. Jiménez, C. Palazuelos, I. Villanueva, *Sampling quantum nonlocal correlations with high probability*, Commun. Math. Phys. 344: 141-154 (2016).
- [13] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association 58 (301), 13-30 (1963).
- [14] T. Jiang, *How Many Entries of A Typical Orthogonal Matrix Can Be Approximated By Independent Normals?*, Ann. Probab. 34(4), 1497-1529 (2006).

- [15] T. Jiang, *Maxima of entries of Haar distributed matrices*, Probability Theory and Related Fields, 131, 121-144 (2005).
- [16] B. Laurent, P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Stat., 28(5), 1302-1338 (2000).
- [17] V. D. Milman, G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed spaces*, Number 1200 in Lectures Notes in Mathematics, Springer Verlag, New York, (1986).
- [18] A. J. Stam, *Limit theorems for uniform distributions on spheres in high- dimensional Euclidean spaces*, J. Appl. Probab. 19 221-228 (1982)
- [19] T. Tao, V. Vu *Random matrices: Universal properties of eigenvectors*, Random Matrices: Theory Appl. 01, 1150001 (2012) and arXiv:1103.2801.
- [20] M. Yor, (1985). *Inégalités de martingales continus arretes a un temps quelconques I*. Lecture Notes in Math. 1118. Springer, Berlin.

**C. E. González-Guillén**

Instituto de Matemática Interdisciplinar, IMI  
Departamento de Matemáticas del  
Área Industrial, E.T.S.I. Industriales, UPM  
28006 Madrid, Spain  
carlos.gguillen@upm.es

**Carlos Palazuelos**

Instituto de Ciencias Matemáticas, ICMAT  
Facultad de Ciencias Matemáticas  
Universidad Complutense de Madrid  
Plaza de Ciencias s/n. 28040, Madrid. Spain  
carlospalazuelos@ucm.es

**I. Villanueva**

Instituto de Matemática Interdisciplinar, IMI  
Facultad de Ciencias Matemáticas  
Universidad Complutense de Madrid  
Plaza de Ciencias s/n. 28040, Madrid. Spain  
ignacioiv@ucm.es